STIC-ILL

From:

Huynh, Phuong N.

Sent: To: Subject:

Thursday, February 13, 2003 5:42 PM STIC-ILL RE: 09/852,976

Please deliver the following:

Immunogenetics 40: 145-149; 1994

Immunogenetics 40: 287-295; 1994

Cancer Immunology, Immunotherapy 51(3): 145-52; 2002

Clin Exp Immunol 110: 509-515; 1997

Thanks, Neon Art unit 1644 Mail 9E12 Tel 308-4844

ORIGINAL PAPER

Shiori Hashimoto · Nicholas Chiorazzi Peter K. Gregersen

The complete sequence of the human CD79b ($\lg\beta/B29$) gene: identification of a conserved exon/intron organization, immunoglobulin-like regulatory regions, and allelic polymorphism

Received: 4 February 1994 / Revised: 29 March 1994

Abstract We determined the complete genomic sequence of the human CD79b (IgB/B29) gene. The CD79b gene product is associated with the membrane immunoglobulin signaling complex which is composed of immunoglobulin (Ig) itself, associated in a noncovalent fashion with CD79b and a second polypeptide chain, CD79a (1ga/mb1). The sequence and exon/intron organization of the human and mouse CD79b genes are highly similar. The gene organization suggests that some variant forms of CD79b may arise by virtue of alternative splicing of mRNA. In addition, a number of conserved regulatory sequences commonly found in Ig genes are present in sequences which flank the human CD79b gene. Some of these sequences are distinct from those found in the CD79a promoter. These differences may explain why transcription of CD79b, but not CD79a, is observed in plasma cells. A new Taq 1 restriction fragment length polymorphism is described that is not associated with any structural polymorphisms of the expressed CD79b polypeptide.

Introduction

The recognition of specific antigen by B lymphocytes depends on the function of the membrane immunoglobulin (mIg) receptor complex, a multimeric complex which, in addition to mIg, includes at least two associated glycoproteins designated CD79a (Igα/mb1) and CD79b (Igβ/ B29) (Reth 1992; Schlossman et al. 1994). These immunoglobulin(Ig)-associated molecules participate in signal transduction and contain highly conserved intracytoplasmic domains which appear to associate with intracellular signalling molecules. The CD79b molecule is of particular interest because it appears that several different forms of CD79b may exist (Friedrich et al. 1993; Hashimoto and co-workers, unpublished results), the molecular basis of which remains undefined. In addition, unlike CD79a (Igal mb1), the transcription of CD79b is maintained in mIg plasma cells (Hashimoto et al. 1993). In order to begin to understand these features of CD79b, we determined the complete sequence of the CD79b gene, and in the course of this work we also defined a new Taq I polymorphism.

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession number L27587

Materials and methods

Isolation of cosmid clones

A chromosome 17-specific cosmid library was obtained from the Reference Library Data Base of the Imperial Cancer Research Fund (London, Great Britain). Filters #1 and #2 (Library number 105, set 12) were screened with a cDNA probe containing the full-length *CD79b* coding sequence, using standard conditions of high-stringency hybridization. Duplicate positive colonies were identified by position on the filter grid, and provided to us by the Reference Library.

DNA sequencing strategy

DNA sequencing was performed by using a Taq DyeDeoxyTM Terminator Cycle Sequencing Kit (ABI, Foster City, CA). Primers specific for portions of the CD79b gene were synthesized (Biosearch 8700; Millipore, Bedford, MA) as needed for sequence analysis by a primer walking strategy. Twenty-two primers were used to complete the se-

S. Hashimoto¹ · N. Chiorazzi · P. K. Gregersen (☑)
Department of Medicine, North Shore University Hospital,
350 Community Drive, Manhasset, NY 11030, USA

N. Chiorazzi - P. K. Gregersen
Department of Medicine, Cornell University Medical College,
New York, NY 10021, USA

¹ Present address: Department of Neurology, Tokyo Women's Medical College, 10 Kawada-cho, Shinjuku-ku, Tokyo, Japan

S. Hashimoto et al.: Genomic sequence of human CD79b (IgB/B29)

s. Hast

quence

reaction

Autom: F13 (5'

the 3^{\prime} .

Southe

Ten 🏨

BRL,

transfe MA). (Strata 15 mi expose probe 15 mi sodiui 0.1 × film a

Resu

Isolı

The (Wo 17-s Libi The hun rato ider vide ider CD to 1 the suc (F1

Sec

The tai or Cl and pe a ex (I re be al si (:

Fig. 1 The complete DNA sequence of the human CD79b (Igβ/B29) gene. Potential regulatory sequences are underlined and labeled above. The exon/intron borders are indicated by a horizontal line, and the adjacent splice donor and acceptor sequences are underlined. The polyadenylation signal is indicated in bold

5. Hashimoto et al., Genomic sequence of manual 95775 (1861)	
TGACCACCGGTGGGGTAAGCACAGACAGAGGGGAGCACAGGCTTCCCCCAGAAGACTGAGAGGCCCCCCAGAGGCATCC	81
'E2-like'	
AGAGGACCC <u>CAGCTTGTC</u> CTGCCCAAGCTGGGCGACCGCCAAACCTTAGCGGCCCAGCTGACAAAAGCCTGCCCT <u>CCCC</u>	162
octamer ets published bri	
tke.	243
	324
ACAGGCTGCAGCCGGTGCAGTTACACGTTTTCCTCCAAGGAGCCTCGGACGTTTCACGGGTTTGGGGTCGGGGACAGA	
THE PROPERTY OF THE PROPERTY O	405
	* 22
· · · · · · · · · · · · · · · · · · ·	486
The second of the second secon	567 648
CCTCCTCTGTCTCTCTCTTTACCGCTGTCTCCTCCTGCCTG	729
SETERTEGECTCAGGAGCCTGTGGCCCCAGTTCCTTCAGGAGGAAGAGAGGCGGAGGGCGCTGCCTGGGAGTCGGGAGGGG AGAGTTAGCAAGCAGGACCGTGGAAGAACTAGGCTCAGGAGGAAGAGAGGCTGAGGGCGGTGCCTGGGAGTCGGGAGTCGGGGCT	810
AGACTTAGCAAGCAGGACCGTGGAAGATCGAAGATCCAAATCTAGTCCCCTCCTCTGGGAGGTGAAGATCCTCGGGCT BAGGCTGTCCCAACCGGACCCTGGAAGATCGAAGATCCAAATCTGGTCCCCTCCTCTGGGAGGTGAAGATCCTCGGTCTCAA	891
PAGGETGTCCCAACCGGACCCTGGAAGATCAGAGATCAACATCAACATCAACATCCCAACATCCCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACATCAACAA	972
The second of th	1053
CCAAGTACTGGGGACAGGGAGGAGGGGGTCTTGAAGCTGCTGTCATGGAAGGAGGAAGTAGCTGCGGGTAGT	1134
GCATGCAGGGGTGGGGCAGCCCAGGCCAGAGGGGGCCCTGACTGGCCCCGCCCACTTCCTCCCACCCA	1215 * 25
EXON 2————————————————————————————————————	1296
	* 39
The second of the second control of the second control of the second of	1377
TO A TOTAL AND A TOTAL PROPERTY OF A TOTAL PRO	1458
	1539
	1620 1701
	1782
TO THE TOTAL PROPERTY OF THE P	1863
	1944
GTGGGCCTCCAGGTCCTGGGGGTGTGGCCTGGGTCAGGCTGTTGTAATGGGCATAGGAAGGA	2025
	2106
CCAGGGAGTCTGCCAGCCAGGCAGGGCAGGGCCTGCTGCTGGGGCTGGCACTCAGGAAGATACCAAGCGGAATGCTGAGCC	2187
TGACCTTGGCACCCACAGGTAGTGCTTGTTCGCGGATCTGGCAGAGCCCACGTTTCATAGCCAGGAAACGGGGCTTCACGG	2268
GS A C S R I W Q S P R F I A R K R G F T V	• 61
TGAAAATGCACTGCTACATGAACAGGGCCTCCGGCAATGTGAGCTGGCTCTGGAAGCAGGAGATGGACGAGAATCCCCAGC	-
K M H C Y M N S A S G N V S W L W K G L H D C C C C C C C C C C C C C C C C C C	* 88 2430 *115
EXON 3	* 88 2430 *115 2511
K M H C Y M N S A S G N V S W L W K G L III EXON 3 AGCTGAAGCTGGAAAAGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTCTACCAGGGCTGCGGCACAGAGCTGCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V	* 88 2430 *115 2511 *143 2593
K M H C Y M N S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N C S A S A S G N V S W L W K G L M D S A S G N C S A S A S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W K G G L M D S G I R F E S Q N E S L A T L T I Q G I R F E S Q N E S L A T L T I Q G I R F E S G N V S W L W R G I R S G	* 88 2430 *115 2511 *142 2593 *14
K M H C Y M N S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N V S W L W K G L M D S A S G N C S G N C S G N E S L A T L T I Q G I R F E S Q N E S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S L A T L T I Q G I R F E S Q N G S C G T G C G G C G G C G G G G G G G G G	* 88 2430 *115 2511 *142 259: *14-267
K M H C Y M N S A S G N V S W L W K G L M S L M S L M S G N V S W L W K G L M S L M	* 88 2430 *11! 251: *14. 259: *14 267 275 *16
K M H C Y M N S A S G N V S W L W K G L M D C M C S M L W K G L M D C M C M C M C M C M C M C M C M C M	* 88 2430 *11! 251: *14: 259: *14 267 275 *16
K M H C Y M N S A S G N V S W L W K G L N T L T S E V Y Q G L R F E L R V L T L T L T L T L R L R L R V L T L R L R L R L R L R L R L R L R L R	* 88 2430 *11! 251: *14. 259: *14 267 275 *16 283 *18
K M H C Y M N S A S G N V S W L W K G L II V P I F L L L D K AGCTGAAGCTGGAAAAGGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTCTACCAGGGCTCCGGCAGAGGTGCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V TCATGGGTGTGTGCCAGGTGGGAGGCCTCCTGTAGCTGTCCGCCCCACCCTAGCCCGGCCTGCCT	* 88 2430 *11: 251 *14 259 *14 267 275 *16 283 *18 291
K M H C Y M N S A S G N V S W L W A G L H X A	* 88 2430 *11! 251: *14: 259: *14 267 275 *16 283 *18 291 299
K M H C Y M N S A S G N V S W L W A G L II O K D EXON 3 ACCTGAAGCTGGAAAAGGGCCGCATGGAAGAGGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTCTACCAGGGCTCCGGCACAGAGCTCCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V TCATGGGTGTGTGCCAGGTGGCTGGGACGCTCCTGTAGCTGTCCGCCCCACCCTAGCCCGGCCTGCCT	* 88 2430 *11! 251: *14. 259: *14 267 275 *16 283 *18 291 299 307
K M H C Y M N S A S G N V S W L W A G L III O R F E EXON 3 AGCIGAAGCTGGAAAAGGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E E S Q N E S L A T L T I Q G I R F E L R V Y Q G C G T E	* 888 2433 **********************************
K M H C Y M N S A S G N V S W L W K G L N C S C N C S W L W K G L N C S C S C S C S C S C S C S C S C S C	243(2431) 243(24
K M H C Y M N S A S G N V S W L W K G L M C L M	* 888 2430 *11! 251:*14. 259:*14 267 275 *16 283 *18 299 307 315 *19 324
K M H C Y M N S A S G N V S W L W K G L N C L N S K S G N V S W L W K G L N C L N C L N C L N C L N S A S G N V S W L W K G L N K G L N C L N C L N N S A S G N V S W L W K G L N K G L N C L N N S A S G N V S W L W K G L N K G L N C L N C L N N S A S G N V S W L W K G L N K G L N C	* 88 243(2 * 11 * 11 * 12 * 12 * 12 * 12 * 12 *
K M H C Y M N S A S G N V S W L W K G L M C L M T L T I Q G I R F E AGCTGAAGCTGGAAAGGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGGAGTGCAACAACACCTCGGAGGTCTACCAGGGCTCCGGCACGAGGTCCGGA D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V TCATGGGTGTGTGCCAGGTGGCTGGGACGCTCCTGTAGCTGTCCGCCCCACCCTAGCCCTGGCCTGCCT	* 888 2430 *111: 2511 *14: 259: *14 267 275 *16 283 *182 291 307 315 *19 332 *2.
K M H C Y M N S A S G N V S W L W K G L M C L M T L T I Q G I R F E L K L E K G R M E E S Q N E S L A T L T I Q G I R F E L R V L T L R D G G T T T L T L R D G G T T T L R D G G T T T T L R D G G T T T T L R D G G T T T T L R D G G T T T T L R D G G T T T T L R D G G T T T T L R D G G T T T T T T T T T T T T T T T T T	* 888 2430 *11! 2511 *14: 259: *14 267. 275 *16 283 *188 291 299 307 315 *19 333: *2: 344
K M H C Y M N S A S G N V S W L W K G L N C L N S K S G N V S W L W K G L N C L N C L N C L N N S A S G N V S W L W K G L N K G L N N S A S G N V S W L W K G L N K G L N N S A S G N V S W L W K G L N K G L N N S A S G N V S W L W K G L N K G L N N S A S G N V S W L W K G L N K G L N N S A S G N V S W L W K G L N N F S E V Y Q G C G I R F E L K G R M E E S Q N E S L A T L T L T L Q G I R F E E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTTACCAGGGCTGCGGCACAGAGGTGCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V M CATGOGGTGTGTGCCAGGTGGGAGGAGCTGCGGCTGCCTTGTGGGCCCCCCCC	* 888 24300 *115 25111 *144 2592 *156 *16 283 *188 291 299 307 315 *199 332 *21
K M H C Y M N S A S G N V S W D W A G D N D S A S G N V S W D W A D W A G D N D S A S G N V S W D W A G D N D D S A S G N V S W D W A D W A G D N D D S A S G N V S W D W A D W A D D N D S A S G N V S W D W A D W A D D N D S A S G N V S W D W D W A D D N D S A S G N V S W D W D W D D N D D S A D D D N D D N D D N D D D D D D D D D	* 888 2430 *119 2511 *144 257: *16 283 *18 291 299 307 315 *24 *19 3324 *21 344 *2:
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	* 888 2430 *115 2511 *144 257: *16 283 *18 291 307 3155 *292 344 *292 344 *293 333 *213 *213 *213 *213 *213
K M H C Y M N S A S G N V S W D W A G D N D G D N F F E EXON 3 AGCTGAAGCTGGAAAGGGCCGCATGGAAGAGGTCCCAGAACGAATCTCTGGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L A T L T I Q G I R F E E AGGACAATGGCATCTCTCTCTGCTGAGGAGGTGCAACAACACCTCGGAGGTTCTACCAGGGCTGCGGCACAGAGGTGCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V Y C A T L T I Q G I R F E E C R V Y Q G C G T E L R V Y Q G C G T T T C T T T C T T C T T C T C T	* 88 2430 *115 2511 *142 2592 *14. 267: *16 283 *18 291 307 315 *199 332 *21 340 *22 344 355
K M H C Y M N S A S G N V S W L W K G L N L T L P S L N T L T L Q G I R F E AGCTGAAGCTGGAAAAGGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L N T L T L Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTTCACCAGGGCTCCGCACAGAGCTCCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V	* 888 24300 *115 25111 *142 25922 *144 2673 2754 *166 2833 *188 2911 2999 3027 3155 *324 *199 3322 *21 344 356 366
K M H C Y M N S A S G N V S W L W K G L L T L T L Q G I R F E L K L E K G R M E E S Q N E S L A T L T I Q G I R F E L K L E K G R M E E S Q N E S L A T L T I Q G I R F E L K L E K G R M E E S Q N E S L A T L T I Q G I R F E L R V Y Q G C G T T E L R V Y Q G C G T T T T T T T T T T T T T T T T T	2349 * 88 2430 *115 2511 *142 2592 *144 2673 2754 283! *18. 2991 307. 315 *196 324 *19 332 *21 340 *22 346 356 37: 38
K M H C Y M N S A S G N V S W L W K G L N L T L P S L N T L T L Q G I R F E AGCTGAAGCTGGAAAAGGGCCGCATGGAAGAGTCCCAGAACGAATCTCTCGCCACCCTCACCATCCAAGGCATCCGGTTTG L K L E K G R M E E S Q N E S L N T L T L Q G I R F E AGGACAATGGCATCTACTTCTGTCAGCAGAAGTGCAACAACACCTCGGAGGTTCACCAGGGCTCCGCACAGAGCTCCGAG D N G I Y F C Q Q K C N N T S E V Y Q G C G T E L R V	* 888 2430 *115 2511 *142 2599 *14. 267: *16 283 *18 291 307 3155 *190 3322 *21 344 *25 344 355 36 37

162

243

324 405

486

567

648

729

810

891

972

1053 1134

1215

* 25

1296

1620

1701

1782

1863

1944

2025

2106

2187

2268

* 61

2349

* 88

2430

*115

2511

*142

2592

*143

2673

2754

*160

2835

*183

2916

2997

3 0 7 8

3159

*196 3240

* 197

3321

3402

*229

3483

3564

3645

3726

3807

3887

quence analysis of approximately 3800 base pairs (bp). Sequencing reactions were analyzed on an Applied Biosystems (Foster City, CA) Automated Sequenator Model 373 A. The most 3' sequencing primer F13 (5' CGG ACA GGG GAA GTG AAG TG 3') was used to confirm the 3' extent of cosmid clones.

Southern blot analysis

Ten µg of genomic DNA were digested overnight with Taq 1 (Gibco/BRL, Gaithersberg, MD), electrophoresed through 1% agarose gel, and transferred onto nylon membranes (Micron Separations, Westboro, MA). After cross-linking using the Stratalinker UV cross-linker (Stratagene, La Jolla, CA) for 90 s, the filters were prehybridized for 15 min at 60° C in QuikHyb solution (Stratagene). Filters then were exposed for 1 h at 60° C in QuikHyb solution to a ³² P-labeled, 800 bp probe derived from the 1gβ cDNA, and subsequently washed twice for 15 min at room temperature in 2 × standard sodium citrate (SSC), 0.1% sodium dodecyl sulfate (SDS), followed by a 30 min wash at 65° C in 0.1 × SSC, 0.1% SDS. Filters were exposed overnight on Kodak XAR5 film at -70° C.

Results

Isolation of cosmid clones for CD79b

The CD79b gene has been localized to chromosome 17q23 (Wood et al. 1993). We therefore obtained a chromosome 17-specific cosmid library for screening from the Reference Library Data Base of the Imperial Cancer Research Fund. The cosmid library was screened with a cDNA clone for human CD79b which was previously isolated in this laboratory (Hashimoto et al. 1993). Seven positive clones were identified on the initial screening. These clones were provided to us by the ICRF laboratories and we subsequently identified two cosmids which appeared to contain the entire CD79b gene. These clones were identified by their ability to be amplified by PCR primers specific for the 5' end of the CD79b cDNA sequence (primer pairs F24/R5), and by successful priming of sequence reactions, using a primer (F13) specific for the 3' end of the the cDNA sequence.

Sequence analysis of the CD79b gene

The entire genomic sequence of the CD79b gene was obtained and is shown in Figure 1. The overall intron/exon organization is highly similar to that reported for the mouse CD79b gene (Reth 1992). The CD79b genes of both human and mouse contain 6 exons. Exon 1 encodes a typical leader peptide sequence; this is followed by exon 2 which encodes a short peptide preceeding the Ig-like domain encoded by exon 3. Exon 4 encodes the hydrophobic transmembrane (TM) portion of the CD79b molecule, whereas the coding regions for the intracytoplasmic (CP) domains are divided between exons 5 and 6. As noted previously (Hashimoto et al. 1993), the mouse and human sequences are highly similar in sequence, especially in the TM and CP regions (>95% similarity in predicted amino acid sequence), as well as sharing the conserved exon/intron structure shown

h GACCACCGGTGGGGTAAGCACAGACAGAGGGGAGCA -305 m AGATGA.CCACAGTGGGA..G..GG..TTA.A.A..

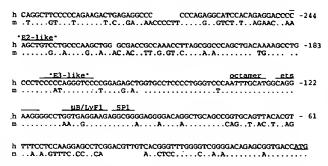


Fig. 2 Sequence comparison of the 5' UT regions of the human and mouse (Hermanson et al. 1988) *CD79b* genes. Conserved regulatory sequences are indicated

here. A 3' untranslated (UT) region of 461 bp separates the stop codon from the polyadenylation signal AATAAA.

In addition to interspecies similarities in the coding region and gene organization, there are also conserved cis-acting regulatory elements which are present in the flanking sequence of human CD79b. A comparison of the human and mouse 5' promoter region is shown in Figure 2. This region contains sequences which are known to be involved in the regulation of Ig genes. Among these are the highly conserved octomer sequence, as well as an ets binding site and an SP1 target sequence. A uB/LyF1 binding region was identified in the mouse and is partially conserved in the human sequence. An additional enhancer (E3-like) element was tentatively identified in the previously reported mouse sequence, and is present in a slightly altered form in the human CD79b promoter. An E2-like sequence is also found in the human sequence, but was not observed in the mouse (Hermanson et al. 1989). Both the human and mouse promoter sequences lack a 5' TATA box. Consistent with this, the mouse and human CD79b transcripts have been shown to initiate at multiple sites downstream of the octomer sequence (Hermanson et al. 1989; Omori and Wall 1993).

Enhancer elements can also be found in the 3''end of Ig genes (Staudt and Lenardo 1991), and for this reason we searched the 3' region of human CD79b for such sequences. We identified four potential sites of E elements, all of which had 7/8 nucleotides identical to the previously described E elements (Staudt and Lenardo 1991). These are indicated in Figure 1.

Identification of RFLPs in the human CD79b gene

We identified an allelic polymorphism associated with the *CD79b* gene after digestion with *Taq* 1. As shown in Figure 3, two bands are observed in homozygous individuals. One band at 2.2 kilobases (kb) is constant in all individuals. A second band reflects a *Taq* 1 polymorphism which is biallelic in the population, with either a 1.8 kb or 2.5 kb



Fig. 3 Southern blot analysis of Taq I-digested genomic DNA from five unrelated individuals. A full-length CD79b cDNA probe was used. Two alleles have been observed in the population. Lane I shows an individual homozygous for the 2.5 kb allele, while lanes 2-4 show individuals homozygous for the 1.8 kb allele. The individual in lane 5 is heterozygous. All subjects contain a constant, and relatively weak, band at 2.2 kb

Table 1 Distribution of Igβ Taq 1 RFLPs in various ethnic groups.

	Caucasian $(n = 20)$	Black $(n = 20)$	Oriental $(n = 20)$	Total $(n = 60)$
2.5/2.5*	3	1	4	8
2.5/1.8	8	3	9	20
1.8/1.8	9	16	7	32

An asterisk indicates that Southern blot analysis of Taq I-digested genomic DNA, using a full-length CD79b cDNA probe, identified two polymorphic fragments of 2.5 kb and 1.8 kb

band present. Heterozygotes (lane 5, Figure 3) contain both of these bands, as well as the constant 2.2 kb band.

We analyzed 60 unrelated individuals for the presence of the Taq 1 RFLP. Subjects from various ethnic groups were examined, and Table 1 shows the results of this study broken down by ethnic group. All three ethnic populations appear to be in Hardy Weinberg equilibrium (null hypothesis not rejected by goodness of fit test).

Discussion

In this study we determined the complete sequence of the human CD79b (Igβ/B29) gene, and the sequence reveals a remarkable degree of similarity between the mouse and human homologues. These similarities relate to both sequence conservation within the coding region, as well as preservation of the overall exon/intron organization. The gene organization is of some interest for understanding the mechanisms by which variant forms of the CD79b molecule may be produced. It has recently been reported that a truncated form of mouse CD79b protein exists in which the C terminal portion of the cytoplasmic tail is deleted (Friedrich et al. 1993). The exact extent of this deletion has not been determined, nor has this been reported as yet in human CD79b.

Nevertheless, the exon/intron structure of the gene suggests that a possible mechanism for this deletion might be alternative splicing at the 3' end of the gene, such that exon 6 is removed from some transcripts. In addition, using an RNAse protection assay, we observed truncated transcripts of the human CD79b in which a large segment of the extracellular portion of the molecule had been deleted (unpublished results). Interestingly, this deletion corresponds exactly to the removal of exon 3, again suggesting

alternative splicing as a mechanism for this variant transcript. The functional significance of these variant forms of *CD79b* has not yet been established.

A second interesting feature of the Ig-associated signaling molecules is their distinct pattern of transcriptional regulation. The CD79b gene is transcriptionally active throughout all stages of B-cell differentiation (Hermanson et al. 1989). Thus, CD79b is expressed with approximately the same timing as the onset of Ig heavy chain rearrangement, an event associated with Ig gene transcription (Blackwell et al. 1986). Consistent with this pattern of expression, the CD79b gene contains regulatory elements commonly found in Ig enhancers. Some of these elements. such as the octomer motif, are identical in the mouse and human CD79b sequences (see Figure 2), as well as in the Ig genes (Staudt and Lenardo 1991). This sequence is known to form a binding site for transcription factors Octl and Oct2. Other putative enhancer elements also exhibit similarity to the previously defined Ig enhancer sequences. Among these are the E3-like and E2-like sequences present in the human CD79b promoter region. It has recently been shown that the E3-like sequence in mouse Ig \beta binds nuclear extracts in gel shift assays (Omori and Wall 1993). In addition to these 5' enhancer elements, we have also identified candidate enhancer sequences in the 3' end of the gene, similar to the location of such enhancers in Ig genes. Four such regions are shown in Figure 1, each of which matches the published Ig enhancer elements in 7/8 nucleotides. Further studies will be required to determine whether these putative regulatory motifs are actually involved in the regulation of CD79b transcription.

As described previously, the major function of the CD79b gene product is in the formation of a membrane signalling complex associated with Ig at the surface of B cells (Reth 1992). At a minimum, this complex consists of Ig itself associated in a noncovalent fashion with CD79b and a second polypeptide chain, CD79a (Igc/mb1). As we have reported previously, the CD79b gene is transcribed at high levels in plasma cells (Hashimoto et al. 1993). In contrast, the CD79a gene is not transcribed in plasma cells (Ha et al. 1992), consistent with the absence of surface Ig in these cells. The CD79a gene contains a promoter region which bears partial similarity to that seen in CD79b; both the ets and SP1 binding sites are present in the CD79a promoter, which also lacks a TATA box. However, no E3-like element is present in CD79a. In addition, a putative cis-acting sequence has been recently identified in CD79a which is the target for a novel DNA binding protein, BlyF (Feldhaus et al. 1992). BLyF is not present in plasma cells, and this may partially account for the lack of transcription of CD79a in plasma cells. Interestingly, a BlyF binding motif is not present in the CD79b 5' promoter region, again emphasizing the distinct patterns of regulation for these two genes. Given the absence of membrane Ig in plasma cells, it is at present unclear what function CD79b has at this stage of B-cell differentiation.

Wood and co-workers (1993) have recently shown that the human *CD79b* gene maps specifically to 17q23. This is of interest in view of the fact that translocations involving 17q23 have been observed in some cases of lymphocytic leukemia (Mitelman et al. 1990). It is tempting to speculate that the presence of numerous enhancer elements for B-cell transcription within the *CD79b* gene might enhance the expression of oncogenes translocated into this region. However, to our knowledge, a detailed molecular analysis of these translocations has not been performed.

Finally, we identified a new Taq 1 polymorphism of the CD79b gene. As shown in Table 1, a preliminary analysis of gene frequencies within various ethnic groups shows a relative preponderance of the 1.8 kb Taq 1 allele with an overall heterozygosity (H) of 0.33; the distribution of genotypes is consistent with Hardy-Weinberg equilibrium in all three ethnic groups. We did not detect any coding region polymorphisms associated with this Taq 1 RFLP, although we have not yet searched for other allelic polymorphisms in the flanking regulatory regions.

Acknowledgments This work was supported in part by US PHS grants from the National Institutes of Health (AI35984 and AI10811) and by the Richard and Nancy Leeds Fund of the Department of Medicine, North Shore University Hospital.

References

Fingnoite at is ng

- Blackwell, T. K., Moore, M. W., Yancopoulos, G. D., Suh, H., Lutzker, S., Selsing, E., and Alt, F. W. Recombination between immuno-globulin variable region gene segments is enhanced by transcription. *Nature* 324: 585-589, 1986
- Feldhaus, A. L., Mbangkollo, D., Arvin K. L., Klug, C. A. and Singh,
 H. BlyF, a novel cell-type- and stage-specific regulator of the
 B-lymphocyte gene mb-1. Mol Cell Biol 12: 1126-1133, 1992

- Friedrich, R. J., Campbell, K. S., and Cambier, J. C. The γ subunit of the B cell antigen-receptor complex is a C-terminally truncated product of the B29 gene. J Immunol 150: 2814-2822, 1993
- Ha, H., Kubagawa, H., and Burrows, P. D. Molecular cloning and expression pattern of a human gene homologous to the murine mb-1 gene. J Immunol 148: 1526-1531, 1992
- Hashimoto, S., Gregersen, P. K., and Chiorazzi, N. The human Ig-β cDNA sequence, a homologue of murine B29, is identical in B cell and plasma cell lines producing all the human Ig isotypes. J Immunol 150: 491-498, 1993
- Hermanson, G. G., Eisenberg, D., Kincade, P. W., and Wall, R. B29: A member of the immunoglobulin gene superfamily exclusively expressed on B lineage cells. *Proc Natl Acad Sci USA 85*: 6890-6893, 1988
- Hermanson, G. G., Briskin, M., Sigman, D., and Wall, R. Immunoglobulin enhancer and promoter motifs 5' of the B29 B-cell-specific gene. Proc Natl Acad Sci USA 86: 7341-7345, 1989
- Mitelman, F., Kaneko, Y., and Trent, J. M. Report of the committee on chromosome changes in neoplasia. Cytogenet Cell Genet 55: 358-386, 1990
- Omori, S. A. and Wall, R. Multiple motifs regulate the B-cell-specific promoter of the B29 gene. *Proc Natl Acad Sci USA 90*: 11723-11727, 1993
- Reth, M. Antigen receptors on B lymphocytes. Annu Rev Immunol 10: 97-121, 1992
- Schlossman, S. F., Boumsell, L., Gilks, W., Harlan, J. M., Kishimoto,
 T., Morimoto, C., Ritz, J., Shaw, S., Silverstein, R. L., Springer,
 T. A., Tedder, T. F., and Todd, R. F. CD antigens 1993. J Immunol 152: 1-2, 1994
- Staudt, L., and Lenardo, M. Immunoglobulin gene transcription. Annu Rev Immunol 9: 373-398, 1991
- Wood, W. J., Jr., Thompson, A. A., Korenberg, J., Chen, X.-N., May, W., Wall, R., and Denny, C. T. Isolation and chromosomal mapping of the human immunoglobulin-associated B29 gene (IGB). Genomics 16: 187-192, 1993